

1 CLAIMS

What is claimed is:

- A method for clustering data comprising:

 - (a) receiving a plurality of data points for clustering;
 - (b) receiving a size parameter for specifying the number of data points to be moved at one time;
 - (c) clustering the data points by using the size parameter to generate clustered results;
 - (d) determining whether the clustered results are satisfactory;
 - (e) when the clustered results are satisfactory, stop clustering;
 - (f) otherwise when the clustered results are not satisfactory, revise the size parameter, perform clustering based on the revised size parameter and the clustered results, and proceed to step (d).

2. The method of claim 1 wherein step (c) further comprises:

(c1) evaluating subsets of data points in each cluster for moving into every other cluster by using a predetermined metric; wherein the number of data points in the subset is specified by the size parameter.

- 18
19 3. The method of claim 2 wherein step (c1) further comprises:
20 (c1_1) determining a geometric center of the subset of data points being evaluated for a
21 move;
22 (c1_2) using the geometric center of the subset of data points in the predetermined metric
23 to generate a value.

- 24
25 4. The method of claim 3 wherein step (c1) further comprises:
26 (c1_3) determining whether the value is greater than zero;
27 (c1_4) when the value is greater than zero, moving the subset of data points from a
28 Move_From cluster to a Move_To cluster;

29 (c1_5) when the value is not greater than zero, determining if there are more subsets to
30 evaluate;
31 (c1_6) when there are more subsets to evaluate, proceeding to step (c1);
32 (c1_7) when there are no more subsets to evaluate, determining whether any point has
33 moved;
34 (c1_8) when a point has moved, proceeding to step (c1); and
35 (c1_9) when no point has moved, stopping the processing.

36
37 5. The method of claim 4 wherein each data point has a membership with one cluster;
38 wherein step (c1_4) further comprises:

39 simultaneously updating the membership of at least two data points from the membership
40 of the Move_From cluster to the membership of the Move_To cluster.

41
42 6. The method of claim 4 wherein step (c1_4) further comprises:
43 updating the count of the Move_From cluster;
44 updating the center of the Move_From cluster;
45 updating the count of the Move_To cluster; and
46 updating the center of the Move_To cluster.

47
48 7. The method of claim 1 wherein revising the size parameter of step (f) further comprises:
49 (f_1) decreasing the size parameter.

50
51 8. The method of claim 1 wherein step (d) further comprises:
52 (d_1) employing a predetermined metric for determining whether the clustered results are
53 satisfactory; wherein the predetermined metric includes a geometric center of the
54 subset of points that are being evaluated for move.

55
56 9. The method of claim 8 wherein the predetermined metric includes the following
57 expression:

58 where U is the subset of data points being evaluated for the move, $|U|$ is the size of U
59 that is specified by the size parameter, m_{oo} is the geometric center of U , M_i and m_j
60 are the centers of the clusters and n_i and n_j are the counts of the clusters.

61
62 10. The method of claim 1 wherein the clustering method is utilized in one of a data mining
63 application, customer segmentation application, document categorization application, scientific
64 data analysis application, data compression application, vector quantization application, and
65 image processing application.

66
67 11. A clustering system comprising:
68 (a) a source of data points to be clustered; and
69 (b) an aggregated clustering module for moving at least two data points at one time
70 between a Move_From cluster and a Move_To cluster;
71 wherein the aggregated clustering module includes
72 a move determination unit for evaluating the move of subsets of data points from each
73 cluster to every other cluster and determining when such a move should be
74 performed; and
75 an aggregated move unit coupled to the move determination unit that updates a
76 Move_From count, Move_From center, a Move_To count, and a Move_To
77 center.
78

79 12. The clustering system of claim 11 wherein the aggregated clustering module further
80 comprises:
81 (a) a first input for receiving the data points;
82 (b) a second input for receiving initial center points;
83 (c) a third input for receiving a number of points to move at one time;
84 (d) a parameter for storing the center point associated with each cluster
85 (e) a parameter for storing the count of data points associated with each cluster;

86 wherein the data points, center points and counts, are utilized by the move determination
87 unit for move evaluation and determination and by the aggregated move unit for
88 count update and center update.

89

90 13. The clustering system of claim 11 wherein the move determination unit further
91 comprises:

92 a geometric center determination unit for determining the geometric center of a current
93 subset of data points and providing the geometric center to the move
94 determination unit for move evaluation and move determination.

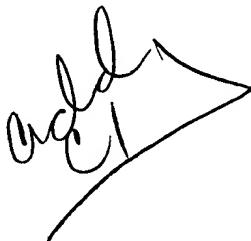
95

96 14. The clustering system of claim 13 wherein the move determination unit further
97 comprises:

98 a move evaluation mechanism for employing a predetermined metric for move
99 evaluation; wherein the predetermined metric includes the geometric center of a
100 current subset of data points.

101
102 15. The clustering system of claim 11 is configured for one of a data mining application,
103 customer segmentation application, document categorization application, scientific data analysis
104 application, data compression application, vector quantization application, and image processing
105 application.

106

A handwritten signature in black ink, appearing to read "Mark A. Johnson". The signature is written in a cursive style with a large, stylized "M" at the beginning. It is positioned at the bottom left of the page, below the text of claim 15.